

## Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors

G. DI CAPUA,<sup>a,b</sup> M. KRETSCHMER,<sup>a</sup> J. RUNGE,<sup>c</sup> A. ALESSANDRI,<sup>d</sup> R. V. DONNER,<sup>a,e</sup>  
B. VAN DEN HURK,<sup>b,g</sup> R. VELLORE,<sup>f</sup> R. KRISHNAN,<sup>f</sup> AND D. COUMOU<sup>a,b</sup>

<sup>a</sup> Potsdam Institute for Climate Impact Research, Potsdam, Germany

<sup>b</sup> Institute for Environmental Studies, VU University of Amsterdam, Amsterdam, Netherlands

<sup>c</sup> Institute of Data Science, German Aerospace Center, Jena, Germany

<sup>d</sup> Royal Netherlands Meteorological Institute, De Bilt, Netherlands

<sup>e</sup> Magdeburg-Stendal University of Applied Sciences, Magdeburg, Germany

<sup>f</sup> Indian Institute for Tropical Meteorology, Pune, India

<sup>g</sup> Deltares, Delft, Netherlands

(Manuscript received 11 January 2019, in final form 15 July 2019)

### ABSTRACT

Skillful forecasts of the Indian summer monsoon rainfall (ISMR) at long lead times (4–5 months in advance) pose great challenges due to strong internal variability of the monsoon system and nonstationarity of climatic drivers. Here, we use an advanced causal discovery algorithm coupled with a response-guided detection step to detect low-frequency, remote processes that provide sources of predictability for the ISMR. The algorithm identifies causal precursors without any a priori assumptions, apart from the selected variables and lead times. Using these causal precursors, a statistical hindcast model is formulated to predict seasonal ISMR that yields valuable skill with correlation coefficient (CC)  $\sim 0.8$  at a 4-month lead time. The causal precursors identified are generally in agreement with statistical predictors conventionally used by the India Meteorological Department (IMD); however, our methodology provides precursors that are automatically updated, providing emerging new patterns. Analyzing ENSO-positive and ENSO-negative years separately helps to identify the different mechanisms at play during different years and may help to understand the strong nonstationarity of ISMR precursors over time. We construct operational forecasts for both shorter (2-month) and longer (4-month) lead times and show significant skill over the 1981–2004 period (CC  $\sim 0.4$ ) for both lead times, comparable with that of IMD predictions (CC  $\sim 0.3$ ). Our method is objective and automatized and can be trained for specific regions and time scales that are of interest to stakeholders, providing the potential to improve seasonal ISMR forecasts.

### 1. Introduction

The Indian summer monsoon (ISM) is a key climate feature critical for Indian society, economy, and ecosystems. The Indian summer monsoon rainfall (ISMR), that is, all-India rainfall averaged over the summer season [June–September (JJAS)], provides about 75% of the total annual rainfall. Thus, enhanced or reduced ISMR strongly affects agriculture output and Indian economy (Gadgil and Rupa Kumar 2006). Moreover,

extremes such as droughts or floods, can seriously affect society and everyday life. Improving ISMR forecasts at sufficiently long lead times would help to plan effective water management strategies, improve flood or drought protection programs, and prevent humanitarian crises.

The complexity and strong internal variability of the ISM circulation system make skillful seasonal forecast challenging (Goswami and Xavier 2005). At seasonal scale, both tropical and northern midlatitude drivers of the ISMR have been identified (Krishna Kumar et al. 1999; Robock et al. 2003; Fasullo 2004; Rajeevan et al. 2007). El Niño–Southern Oscillation (ENSO) affects the ISM strength via the horizontal displacement of the Walker circulation. During El Niño years, the convection is shifted toward the eastern central Pacific, weakening the ISM (Krishna Kumar et al. 1999) and making severe droughts more likely. However, the

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-19-0002.s1>.

Corresponding author: Giorgia Di Capua, [dicapua@pik-potsdam.de](mailto:dicapua@pik-potsdam.de)

correlation between ENSO and ISMR is nonstationary (Robock et al. 2003), and the exact physical mechanisms are not well understood (Kripalani et al. 2001). The intrinsic decadal variability of ISMR can further enhance the impact of ENSO on the occurrence of dry and wet spells, that is, El Niño's drying effect is stronger during drier decades (Kripalani et al. 2001). The correlation between ENSO and the ISMR has weakened over recent decades (Krishna Kumar et al. 1999; Cherchi and Navarra 2013), nevertheless, statistical seasonal prediction models for the ISMR demonstrate some predictability originating from ENSO (Mujumdar et al. 2007; Shukla et al. 2011).

The snow–monsoon mechanism proposes that altered snow cover conditions over Eurasia can modify surface variable characteristics and the radiative balance via snow–albedo and soil moisture effects. Increased snow cover might decrease the land–ocean temperature gradient, reducing the ISMR (Bhanu Kumar 1988; Bamzai and Shukla 1999; Kripalani et al. 2001; Dash et al. 2004) and cause large-scale circulation changes in the midlatitude in winter and spring, which in turn affect the ISMR (Dash et al. 2005, 2006). This relationship is neither stationary nor spatially homogeneous (Bamzai and Shukla 1999; Kripalani et al. 2001; Robock et al. 2003) and depends on ENSO (Fasullo 2004). However, model experiments show that snow conditions alone can significantly affect the ISM circulation (Ferranti and Molteni 1999; Peings and Douville 2010; Turner and Slingo 2011).

Seasonal forecasts of ISMR from atmospheric general circulation models (AGCM) are usually initialized on 1 May and their skill is fairly modest (Weisheimer and Palmer 2014). Dynamical models show that prescribing tropical sea surface temperature (SST) patterns leads to erroneous correlation patterns between the SST and the rainfall, while coupling AGCM to oceanic models helps to improve the represented teleconnections (Wang et al. 2005). Seasonal forecasts from the Climate Historical Forecast Project show an improvement in forecasts skill both for single models and multimodel ensembles, with correlations between the observed and forecasted rainfall that go up to  $\sim 0.4$  over the land-only Indian region, although large model biases remain (Jain et al. 2019).

To complement dynamical forecasts, the India Meteorological Department (IMD) uses a long-range statistical forecasting model for the ISMR based on correlations between different precursors of the atmosphere–ocean system (Rajeevan et al. 2004, 2007; Kumar 2012). The IMD operational forecasting system provides ISMR forecasts in April, June, and July, based on three sets of 6–9 precursors (Kumar 2012). This system has undergone major changes in 2003 (Rajeevan et al. 2004), 2007

(Rajeevan et al. 2007), and 2012 (Kumar 2012), following its failure in forecasting critical years such as 2002 and 2004. Different techniques have been applied, from multilinear regression models (Rajeevan et al. 2004), to multimodel ensembles (Rajeevan et al. 2007) and neural networks (Kumar 2012). The most recent version of IMD precursors (in 2019) features five precursors, which require data till March and forecasts are provided in April, with an update in June.<sup>1</sup> The correlation between precursors and the ISMR is nonstationary, and therefore the precursors require frequent updates (Rajeevan et al. 2002, 2004).

Anomalies in boundary conditions like snow cover, SST, and soil moisture from both midlatitudes and tropical regions are critical to determine the ISMR (Shukla and Mooley 1987). In addition to the influence of ENSO, other atmospheric and surface conditions from mid- and high latitudes are used. Several precursors, that is, atmospheric and surface fields averaged on a certain spatial domain and usually determined by correlation relationships with the ISMR at a certain lead time, have been used to forecast the ISMR with statistical models (Rajeevan et al. 1998, 2005, 2007). Here, we briefly summarize some of the precursors used by IMD for their statistical forecasts. Some predictability comes from the Eurasian and European regions (Rajeevan et al. 2005). An enhanced pressure gradient between northern and southern Europe in January is linked to stronger ISMR: the North Atlantic Oscillation (NAO) may affect the Eurasian westerly flow and therefore snow cover anomalies over Eurasia (Rajeevan 2002; Goswami et al. 2006). IMD also uses temperature anomalies over the Scandinavian region (Rajeevan et al. 2005), while warm temperature anomalies over the Northern Hemisphere in January have been linked to enhanced ISMR (Sikka 1980; Verma et al. 1985). Positive surface temperature anomalies over Eurasia, north of 60°N, show significant positive correlation with excess ISMR years, while central Asia, around 40°N, sees a reversed relationship (Rajeevan et al. 1998). Negative February–March pressure anomalies over East Asia due to a westward shift of the Aleutian low might alter typhoon tracks during the monsoon season and are linked to deficient ISMR (Saha et al. 1981; Rajeevan et al. 2005). IMD identifies some predictability for the ISMR also from SST in the North Atlantic east of the Gulf of Mexico and sea level pressure (SLP) from the Azores High region (likely linked to the NAO) (Rajeevan et al. 2005, 2007).

---

<sup>1</sup> Further details regarding the latest version of the IMD forecast model are available at the following link: [http://www.imd.gov.in/pages/monsoon\\_main.php?adta=PDF&adtb=2](http://www.imd.gov.in/pages/monsoon_main.php?adta=PDF&adtb=2).

Another IMD precursor is represented by SST over the southeast Indian Ocean during February and March. Here, higher SST may induce enhanced evaporation and thus fuel stronger ISMR (Rajeevan et al. 2002; Terray et al. 2003). Surface winds and warm water volume over the Niño-3.4 region also enter the set of IMD precursors (Rajeevan et al. 2005).

Next to the nonstationary nature of ISMR precursors (Krishna Kumar et al. 1999; Rajeevan 2001), “spurious” correlations overshadowing the actual physical links, overfitting, and decadal variability in the correlation structure also represent challenges for statistical ISMR forecasts (Hastenrath and Greischar 1993). Selection of precursors purely on their correlation with the ISM introduces a significant risk of including noncausal relationships and of overfitting the model (DelSole and Shukla 2009).

Machine learning techniques such as artificial neural networks (ANN) have also been applied to the ISMR forecasting problem, showing promising results for both monthly and seasonal ISMR at 1-month lead time (Sahai et al. 2000; Singh and Borah 2013; Singh 2018). Singh (2018) trains ANN with past values of monthly ISMR time series for the period 1871–1960 and provides forecasts for the period 1961–2014; to forecast all-India rainfall amount for June, monthly all-India rainfall values for the months from January to May are used. The promising results obtained with ANN show the potential of machine learning techniques in improving ISMR forecasts.

Recently, a novel statistical forecast method based on a causal discovery algorithm was introduced, designed to identify *causal* precursors of a variable of interest, thereby limiting the risk of overfitting (Kretschmer et al. 2017). This Response-Guided Causal Precursors Detection (RG-CPD) algorithm combines spatial clustering (Bello et al. 2015) with causal discovery (Runge et al. 2014, 2015a,b). Without requiring an a priori definition of the possible precursors, RG-CPD searches for correlated precursor regions of a variable of interest in multivariate gridded data and then detects *causal* precursors by filtering out spurious links due to common drivers, autocorrelation effects, or indirect links. This approach was shown to be able to extract known physical pathways affecting the stratospheric polar vortex and thereby deliver skillful forecasts of the vortex’s strength at lead times up to two months (Kretschmer et al. 2017).

Here, we apply the RG-CPD scheme to detect causal precursors of ISMR and show that the method can provide skillful hindcasts at 4-month lead time. We analyze the influence of the ENSO background state on casual precursors. The sensitivity of the results to the choice of the observational data is tested by using two

different rainfall datasets. We also test the robustness of our methodology given the nonstationarity of the precursors. Finally, we construct an operational forecast model based on causal precursors and compare the results with existing operational forecasts from IMD, showing that causal discovery techniques have the potential to improve seasonal forecast of ISMR over India.

## 2. Data and methods

### a. Data

We analyze all-India summer (JJAS averaged) monsoon rainfall (ISMR) from the Climate Prediction Center–National Centers for Environmental Prediction (CPC–NCEP) observational gridded ( $0.25^\circ \times 0.25^\circ$ ) global rainfall dataset (in the following briefly referred to as CPC) for the period 1979–2016 (Chen et al. 2008) and from the Rajeevan ( $1^\circ \times 1^\circ$ ) observational gridded rainfall dataset (RAJ) over India for the period 1901–2004 (Rajeevan et al. 2008). Precursor variables are taken from monthly averaged gridded ( $1.5^\circ \times 1.5^\circ$ ) SLP and 2-m surface temperature (T2m) fields from the ERA-Interim (Dee et al. 2011) for the period 1979–2016 and ERA20C reanalyses for the period 1901–2004 (Poli et al. 2016). We choose T2m and SLP because they provide fairly reliable data over the twentieth century and historically have been used by IMD for their statistical forecast (Rajeevan et al. 2005, 2007).

Figure 1 shows the summer (JJAS) 1979–2016 climatology for CPC (Fig. 1a) and 1979–2004 climatology for RAJ (Fig. 1b), the time series of summer mean CPC over the 1979–2016 period and RAJ over the 1951–2004 period (Fig. 1c). IMD operational forecasts<sup>2</sup> compared with IMD observed ISMR are shown in Fig. 1d. The correlation coefficient between CPC and RAJ over the common period (1979–2004) is  $CC = 0.58$ , highlighting some uncertainty in the data. For CPC and RAJ time series, anomalies relative to the mean seasonal climatology are calculated and the data are detrended.

### b. The response-guided causal precursor detection algorithm

RG-CPD is a recently developed algorithm that identifies the *causal precursors* of a *response variable* based on multivariate gridded observational data (Kretschmer et al. 2017). RG-CPD combines a response-guided detection step (Bello et al. 2015) with a causal discovery step (Spirtes et al. 2000; Runge et al. 2012, 2014,

<sup>2</sup> IMD seasonal forecasts can be found at [http://imd pune.gov.in/Clim\\_Pred\\_LRF\\_New/Home/LRF\\_Perform\\_89-2017.html](http://imd pune.gov.in/Clim_Pred_LRF_New/Home/LRF_Perform_89-2017.html).

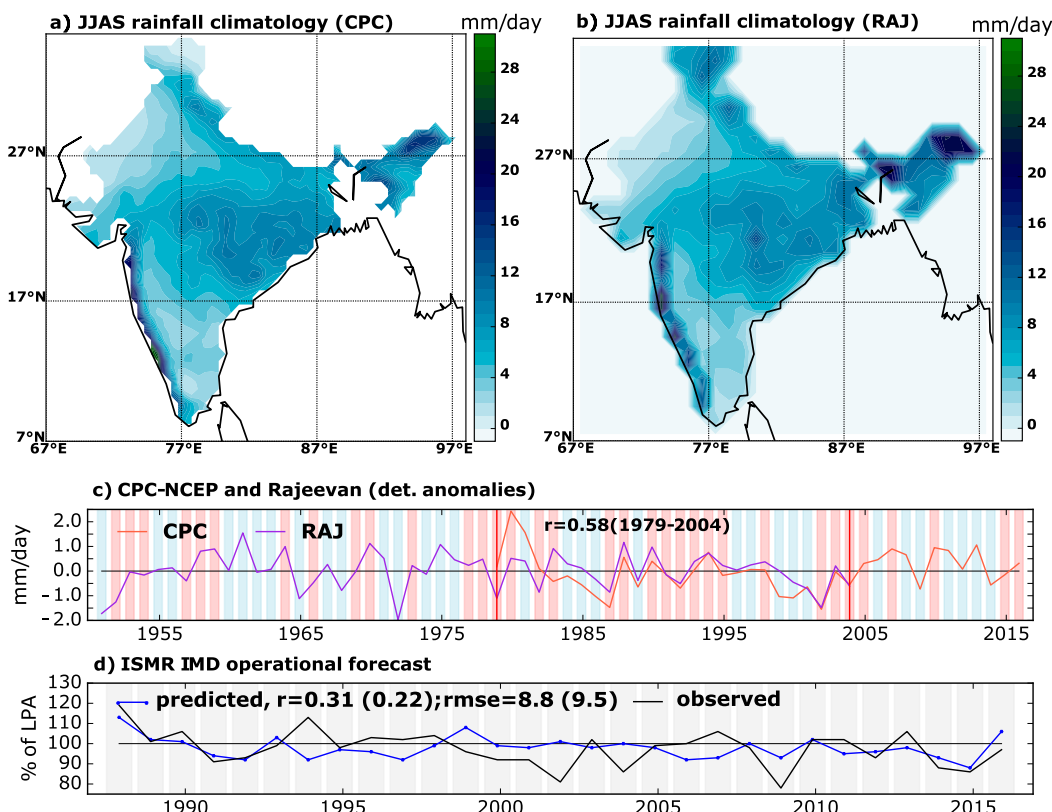


FIG. 1. Rainfall datasets. ISMR (JJAS averaged) climatology over India for (a) CPC–NCEP (CPC) covering 1979–2016 and (b) Rajeevan (RAJ) covering 1951–2004. (c) Detrended JJAS-mean time series for all India rainfall (ISMR) for CPC (red) and RAJ (purple). ENSO-positive (ENSO-negative) years are highlighted by red (blue) background shading. (d) ISMR for 1988–2016 for observed IMD ISMR (black solid line) and the IMD operational forecast (solid blue line), expressed in percent anomaly from the long period (50 years) average (LPA). Correlation and RMSE values are reported for both the 1988–2016 period and for the comparison period 1988–2004 (in parentheses).

2015a,b). Using correlation maps, an initial set of precursor regions is identified in relevant meteorological fields by finding regions that precede changes in the response variable at some lead time (response-guided detection step). In the second step, an adapted version (Runge et al. 2014) of the Peter and Clark (PC) causal discovery algorithm (Spirtes et al. 2000) iteratively tests whether the correlation of a precursor with the response variable can be explained by the influence of any other precursor or combination of precursors. Noncausal precursors due to spurious correlations caused by indirect links, common drivers, or autocorrelation effects are removed and only the causal precursors remain. Note that the term *causal* rests on several assumptions (Spirtes et al. 2000; Runge 2018) and in this context should be understood as *causal relative to the set of analyzed precursors*.

Here, we apply the RG-CPD scheme to identify causal precursors of observed ISMR anomalies (Fig. 1c). We search for causal precursors in the global fields of SLP and T2m at 4- and 5-month lead times (i.e., signatures in February and January respectively). In the first RG-CPD

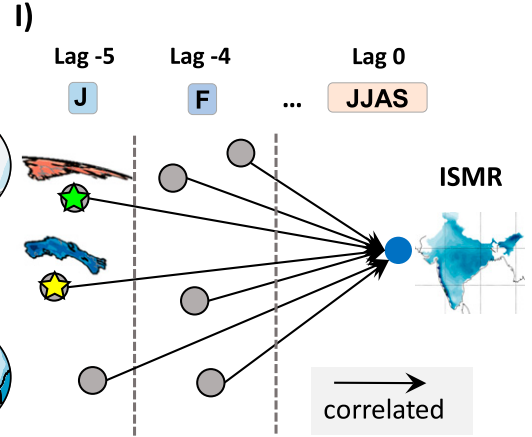
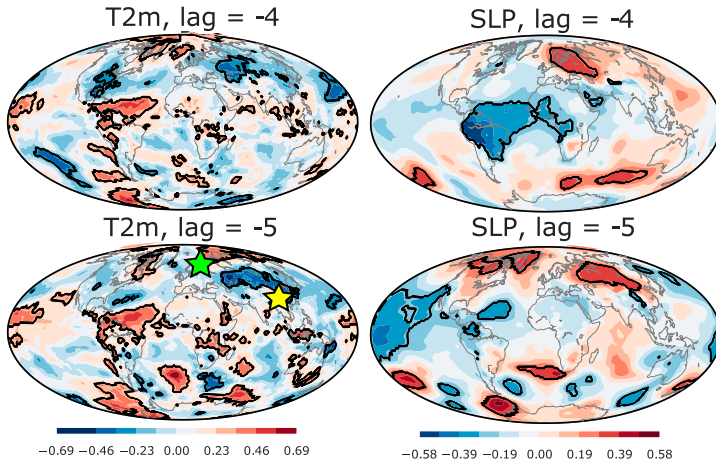
step, the ISMR time series is correlated with monthly SLP or T2m fields in February (F) (4-month lead) and January (J) (5-month lead), see Fig. 2a. In all correlation maps, adjacent grid points with a significant correlation of the same sign at a level of  $\alpha = 0.05$  (accounting for a two-tailed Student's  $t$  test ignoring autocorrelation effects and field significance) are spatially averaged to create single time series: each region is reduced to one single one-dimensional time series, called precursor region (Willink et al. 2017). The set of precursor regions (PR) is defined as

$$\text{PR} = \{T2m_{\tau=-4}^1, \dots, T2m_{\tau=-5}^m, \dots, \\ SLP_{\tau=-4}^1, \dots, SLP_{\tau=-5}^n\},$$

where  $\tau$  is the lag at which the correlation is detected expressed in months and  $m$  and  $n$  represent the total number of identified precursor regions for T2m and SLP, respectively. By design, each element in PR is significantly correlated with the response variable ( $\text{ISMR}_{\tau=0}$ ), forming a dense correlation network [Fig. 2a(I)]. As an



### a) One-point correlation maps



### b) Causal Precursors

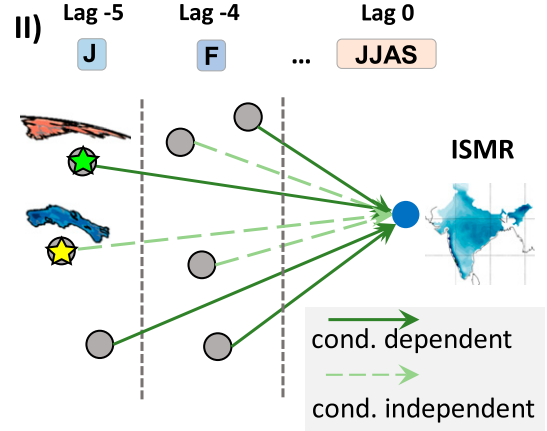
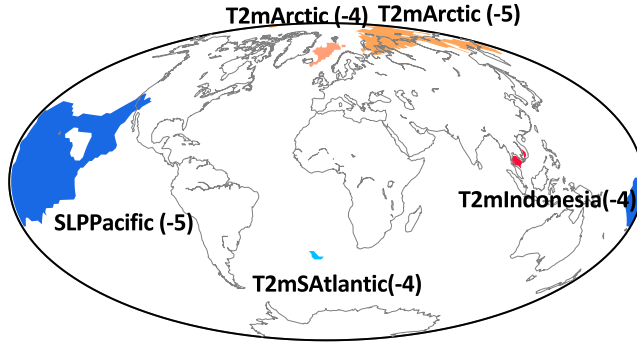


FIG. 2. Causal precursors over the 1979–2016 period. [a(I)], [b(II)] A schematic figure of the RG-CPD algorithm. (a) (left) Correlation maps of T2m and SLP with ISMR at 4- and 5-month lead times. In [a(I)] the first step of RG-CPD that identifies a large number of spatial regions (gray dots) that correlate (black solid lines) with ISMR is shown schematically. (b) (left) The subset of causal precursors, with name and lead time, detected with the second step of RG-CPD algorithm. In [b(II)] this second step is illustrated schematically: causal links are shown by solid green arrows (*conditionally dependent* with ISMR), while spurious links are shown by dashed green arrows (*conditionally independent* with ISMR). The green and yellow stars in [a(I)] and [b(II)] identify the causal precursor  $T2m_{\tau=-5}^{\text{Arctic}}$  and the spuriously correlated  $T2m_{\tau=-5}^{\text{CAsia}}$  regions, respectively.

example, we report the correlation for the regions  $T2m_{\tau=-5}^{58}$  and  $T2m_{\tau=-5}^{56}$  (green and yellow stars in Fig. 2a) and their  $p$  values:

$$\rho = \rho(\text{ISMR}_{\tau=0}, T2m_{\tau=-5}^{56}) = -0.48, \quad p = 0.0004,$$

$$\rho = \rho(\text{ISMR}_{\tau=0}, T2m_{\tau=-5}^{58}) = 0.54, \quad p = 0.002.$$

In the second step, the PR set is searched for causal precursors (causal discovery step). The (adapted) PC algorithm (Tigramite version 2.0, <https://github.com/jakobrunge/tigramite>) tests whether the lagged correlation between each precursor and the response variable can be explained by the influence of other precursors, by calculating partial correlations iterating through

different combinations of precursors. In the first iteration, the partial correlation between ISMR and each element of the PR set is calculated conditioning on each of the remaining elements of the PR. Hence, the partial correlation between the variables  $x$  and  $y$  conditioned on variable  $z$  is computed by first performing a linear regression of  $x$  on  $z$  and of  $y$  on  $z$  and then calculating the correlation between the residuals:

$$\rho = \rho(x, y|z) = \rho[\text{Res}(x), \text{Res}(y)].$$

If the partial correlation between  $x$  and  $y$  is still significant at a certain significance threshold  $\alpha$ ,  $x$  and  $y$  are said to be *conditionally dependent* given variable  $z$ , that is, the correlation between  $x$  and  $y$  cannot be

(exclusively) explained by the influence of variable  $z$ . This process is continued by conditioning on one further condition in each iteration step. After each iteration, precursors whose partial correlation with the ISMR is not significant given a threshold  $\alpha = 0.05$  are removed. The algorithm converges when there are no further variables to condition on.

Continuing with our example and conditioning on  $T2m_{\tau=-5}^{26}$  (precursors are ordered by the strength of the correlation), we have that

$$\begin{aligned}\rho &= \rho(\text{ISMR}_{\tau=0}, T2m_{\tau=-5}^{56} | T2m_{\tau=-5}^{26}) = -0.61, \\ p &= 0.00001.\end{aligned}$$

The partial correlation between  $\text{ISMR}_{\tau=0}$  and  $T2m_{\tau=-5}^{56}$  conditioned on  $T2m_{\tau=-5}^{26}$  is still significant ( $p$  value  $< 0.05$ ), thus  $\text{ISMR}_{\tau=0}$  and  $T2m_{\tau=-5}^{56}$  are conditionally dependent given the variable  $T2m_{\tau=-5}^{26}$ .  $T2m_{\tau=-5}^{56}$  is ultimately filtered out by a combination of two precursors, as follows:

$$\begin{aligned}\rho &= \rho(\text{ISMR}_{\tau=0}, T2m_{\tau=-5}^{56} | T2m_{\tau=-5}^{26}, T2m_{\tau=-4}^2) = -0.30, \\ p &= 0.05.\end{aligned}$$

Thus, the correlation between  $\text{ISMR}_{\tau=0}$  and  $T2m_{\tau=-5}^{56}$  can be eventually explained by the combined effect of  $T2m_{\tau=-5}^{26}$  and  $T2m_{\tau=-4}^2$ . Hence,  $\text{ISMR}_{\tau=0}$  and  $T2m_{\tau=-5}^{56}$  are *conditionally independent*, the correlation with the response variable is spurious and therefore  $T2m_{\tau=-5}^{56}$  is filtered out. Figure 2a(II) shows a schematic of this step. Contrarily, the partial correlation of  $T2m_{\tau=-5}^{58}$  with ISMR is significant even after conditioning on the influence of a range of different combinations of precursors in PR. Thus,  $T2m_{\tau=-5}^{58}$  and ISMR are found to be conditionally dependent and  $T2m_{\tau=-5}^{58}$  is interpreted as a causal precursor of ISMR (see Kretschmer et al. 2016, 2017 for a more detailed description).

### c. Statistical hindcasting and forecasting models

We test the hindcast skill of the causal precursors (identified by applying RG-CPD over the full time period) by performing a multivariate linear regression of the 4- and 5-month lead time causal precursors of ISMR. To assess the performance of the regression model, we divide the time series into training and testing periods containing 25 and 13 years, respectively. To avoid using any information from the predicted year in the regression model, all data processing (i.e., calculating anomalies and detrending the data) is based on the training period only. However, although the training of the regression model is independent of the testing period, it still contains some information from the testing period

as the causal precursors are detected on the full period (training plus testing, see Fig. 3a). We refer to this prediction method as “hindcast.”

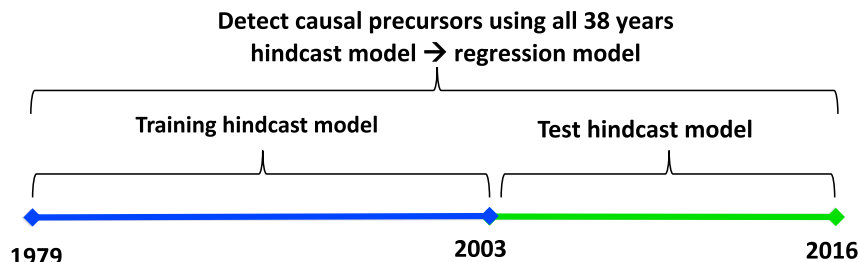
When testing for the influence of ENSO, we hindcast ISMR using leave-one-out cross validation instead of the training–testing period approach. In this case, the regression model is trained by leaving out all data from the year of the summer to be hindcasted, and thus the length of the training period is always equal to the total length of the time series minus 1. For example to predict 1979 ISMR, we train the model with data from the 1980–2016 period and so on. We use the leave-one-out method because the time series for ENSO-positive and ENSO-negative years alone are too short ( $\sim 20$  years each) to be divided into training and testing period.

Next, we build an operational forecast model. For this purpose we apply RG-CPD on a moving window of 30 years (training period), and then forecast the year 31 (testing period), as would be the case in an operational forecast system. This way, no information (including the detection of the causal precursors) from the year to be forecasted enters the forecast model. We refer to this prediction method as an (operational) forecast (see Fig. 3b for schematic visualization).

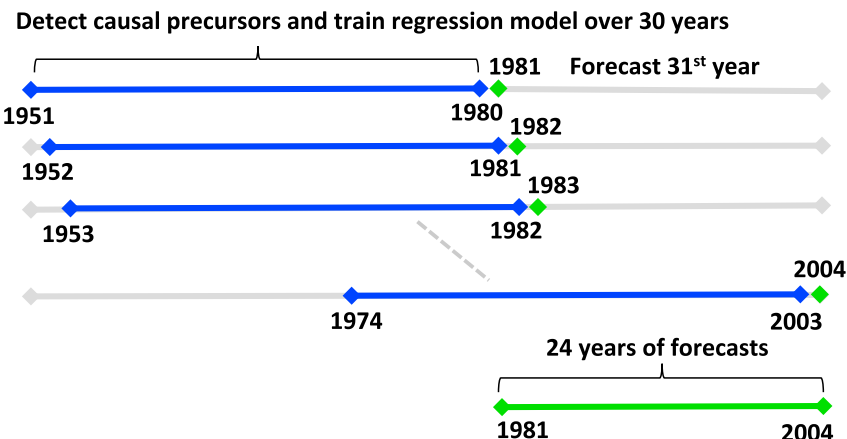
To identify robust causal precursors, we first calculate the correlation maps (first RG-CPD step) over the 30-yr window (Fig. 3c). The choice of using a 30-yr window is motivated by the need to balance the nonstationarity of the precursors (which thus requires a shorter time series) and the need to satisfy the statistical requirements of the causal discovery algorithm (with longer time series preferred). Next, we divide this 30-yr window into five subsets of 24 years by excluding a moving interval of 6 years and then apply the causal discovery step (second RG-CPD step) to each of these five subsets (Fig. 3c). For each 30-yr window, we obtain five sets of causal precursors. To predict year 31, we use *all* identified causal precursors of the five subsets (using overlapping causal precursors just once).

To assess the skill we calculate the receiver operating characteristic (ROC) curve, a common metric to evaluate the skill in predicting predefined observed events, illustrated by the true and false positive rates for different threshold settings of the predicted time series (Wilks 2006). For a given set of observed events, the true positive rate is the number of correctly predicted events (i.e., the number of events in the predicted time series that exceeds a certain threshold), normalized by the number of observed events. Likewise, the false positive rate is the rate of wrongly predicted events. An area under the ROC curve (AUC) of 1.0 means perfect hindcast (or forecast) skill while an AUC of 0.5 means no skill.

### a) Hindcasting



### b) Forecasting



### c) Optimal causal precursor detection for forecasting

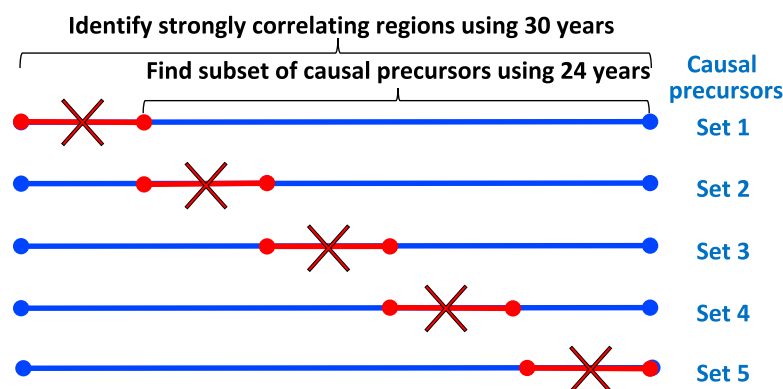


FIG. 3. Summary of employed hind- and forecasting strategies. (a) A schematic of the hindcasting method whereby the causal precursors are identified using all 38 years (1979–2016) and the regression model is trained over the first 25 years (1979–2003) and tested over the last 13 years (2004–16). (b) A schematic of the forecasting method whereby the causal precursors are identified and the regression model is trained using 30 years of data to forecast year 31. (c) A schematic of how a robust set of causal precursors are identified for the forecasting method shown in (b).

## 3. Results

### a. Causal precursors of ISMR

Correlation maps of CPC and SLP/T2m respectively at 4- and 5-month lead times (Fig. 2a), give a set of 112

possible precursor regions in T2m and 28 in SLP (see Table S1 in the online supplemental material). A few features arise from the correlation maps: in general, SLP patterns tend to be larger and smoother than T2m patterns. Here, a negative correlation means that low

pressure anomalies over a certain area during January (5-month lead) or February (4-month lead) are followed by enhanced ISMR. SLP patterns show a large region of negative correlation over the tropical belt at both lead times, centered in the western tropical Atlantic at 4-month lead and over central tropical Pacific at 5-month lead. In general, northern mid- to high latitudes positively correlate with ISMR, with higher-pressure anomalies over the Arctic and midlatitude regions during boreal winter being followed by enhanced ISMR. Relevant T2m patterns are spatially more confined and scattered than SLP precursors. Notably, the T2m correlation maps in the midlatitude show a dipole extending from a positively correlated region over the Arctic toward a negatively correlated region over Eurasia. The tropical belt shows a general tendency toward positive correlations with T2m, with larger centers over the western central Atlantic and central Pacific (partly reflecting the SLP patterns).

After applying the PC algorithm, only five causal precursors remain (Fig. 2b): two T2m regions in the Arctic,  $T2m_{\tau=-5}^{\text{Arctic}}$  at 5-month lead time and  $T2m_{\tau=-4}^{\text{Arctic}}$  at 4-month lead time, a tropical SLP band over the central Pacific at 5-month lead time ( $SLP_{\tau=-5}^{\text{Pacific}}$ ), a small T2m region north of Indonesia at 4-month lead time ( $T2m_{\tau=-4}^{\text{Indonesia}}$ ) and T2m over the southern Atlantic at 4-month lead time ( $T2m_{\tau=-4}^{\text{SAntlantic}}$ ). Thus, more than 95% of strongly correlated regions are removed, showing how pervasive spurious correlations are and how important it is to account for this effect. Causal precursors for 2- and 3-month lead times are shown in Fig. S1.

Next, we test the hypothesis that the causal precursors of CPC depend on the ENSO background state. To study the influence of ENSO on CPC, we divide the 1979–2016 period into two subsets corresponding to positive and negative Niño-3.4 index averaged over the January–April period (here briefly referred to as ENSO-positive and ENSO-negative years, respectively,<sup>3</sup> see Fig. 1c) and rerun RG-CPD for each of these subsets separately. Due to the shortness of the time series, it is not feasible to differentiate between neutral, positive, and negative ENSO phases. In total, we define 21 ENSO-positive and 17 ENSO-negative years (Fig. 1c).

The correlation patterns for ENSO-positive and ENSO-negative years are very different from those shown above (Fig. 4), illustrating that the causal precursors shown before (Fig. 2b) are linked to either positive or negative ENSO conditions. This does not need to imply that the causal precursors resemble classical El Niño or La Niña

patterns. In fact they do not. Rather, the different causal precursors suggest that the ENSO cycle redistributes the background state and thereby creates possibilities for different far-away precursors to dominate. During ENSO-negative years, the Arctic is not a precursor, while during ENSO-positive years its positive T2m correlation with CPC strengthens (from  $\sim 0.5$  to  $\sim 0.7$ ). The negatively correlated SLP region over the Pacific shifts from the tropical Atlantic/Indian Ocean during ENSO-positive years to the Central Pacific/tropical Atlantic during ENSO-negative years. The signal from the tropical North Atlantic intensifies during ENSO-negative years and disappears for ENSO-positive years. During ENSO-positive years, three regions are detected as causal precursors of CPC:  $T2m_{\tau=-4}^{\text{Arctic}}$ , a negatively correlated T2m region over Central Asia ( $T2m_{\tau=-5}^{\text{CAAsia}}$ ) and a positively correlated T2m region close to west Antarctica ( $T2m_{\tau=-4}^{\text{WAntarctica}}$ ) [Fig. 4a(II)]. Two regions are found to be causal precursors of CPC during ENSO-negative years: T2m over the southwest Indian Ocean ( $T2m_{\tau=-5}^{\text{SWIndianOc}}$ ), and T2m over the tropical North Atlantic ( $T2m_{\tau=-5}^{\text{NAtlantic}}$ ) [Fig. 4b(II)].

#### b. Hindcast based on causal precursors

We now test the predictive skill of the identified causal precursors over the full 1979–2016 period. The five precursor regions identified in Fig. 2b are used to build a multiple linear regression hindcast model, using 25 years to train the model and 13 years to test it (Fig. 3a). Figure 5a shows the results for the training (blue line) and testing period (green line), together with the correlation with observed ISMR over the training period ( $CC = 0.88$ ) and over the testing period ( $CC = 0.85$ ). Table S2 reports the regression coefficients for each region of the linear regression hindcast model. The ROC curves for hindcasting dry events (i.e., values below the empirical 30th percentile of all ISMR values) and for wet events (above the 70th percentile) are shown in Figs. 5c and 5d, respectively. The resulting AUC values are 0.94 for dry events and 0.96 for wet events (Figs. 5c,d,  $p$  values  $< 0.05$ ), meaning that the model has good skill in hindcasting CPC at 4-month lead time. Here, the significance of the AUC score has been calculated via shuffling techniques.

The hindcast performed separately for ENSO-positive and ENSO-negative years along with leave-one-out cross-validation is shown in Fig. 5b, together with correlation values for the regression model compared to observations. To hindcast ISMR during ENSO-positive years (shaded in red in Fig. 5b), we use only the causal precursors identified over the ENSO-positive years (shown in Fig. 4a). In the same way, to hindcast ISMR during ENSO-negative years (shaded in blue in Fig. 5b),

<sup>3</sup> This does not imply that a given year exhibited some El Niño or La Niña according to their standard definitions.



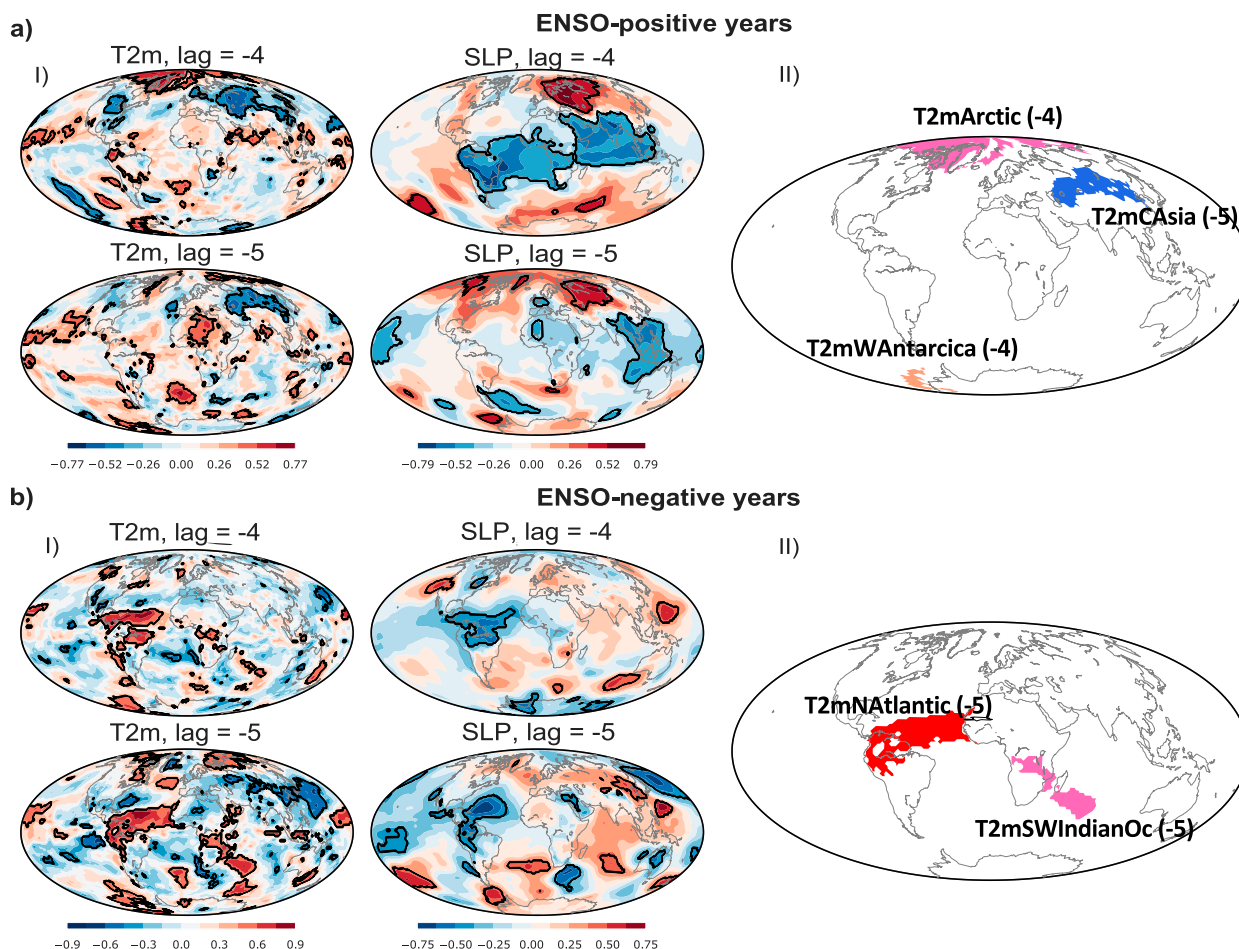


FIG. 4. ENSO influence on causal precursors. [a(I)] Correlation maps between JJAS-mean CPC and (left) T2m or (right) SLP from the ERA-Interim reanalysis during ENSO-positive years. The top row shows the 4-month lead (February) and the bottom row the 5-month lead (January). Regions that are significantly correlated ( $p$  value  $< 0.05$ ) are contoured in solid black lines. [a(II)] The detected subset of causal precursors, identified by name and lead time. [b(I)], [b(II)] As in [a(I)] and [a(II)], but for ENSO-negative years.

we use the causal precursors shown in Fig. 4b. The AUC values obtained for ENSO-positive and ENSO-negative years separately improve slightly with AUC = 0.96 for dry events during ENSO-positive years and AUC = 0.97 for wet events during ENSO-negative years (Figs. 5c,d), showing that considering the ENSO phase might help to improve hindcast skill. Combining the specific hindcast models for ENSO-positive and ENSO-negative into a single hindcast time series (Fig. 5b), we obtain AUC values of 0.93 for dry and 0.92 for wet events. Table S2 reports the regression coefficients for the hindcast model for ENSO-positive and ENSO-negative years.

In addition to ROC analysis, we performed model selection with respect to the given set of precursors by calculating the Akaike information criterion (AIC) for the full 1979–2016 period and for different ENSO phases. AIC is a criterion for selecting the most appropriate among a finite set of models; the model with the lowest

AIC is generally preferred. AIC values are calculated for different combinations of one, four, and five causal precursors identified over the 1979–2016 period. In all three cases, the combination of all five causal precursors exhibits the lowest AIC value with respect to all individual precursors or any combination of four precursors (Table 1). Hence, AIC demonstrates that adding causal precursors leads to better hindcasts without running into overfitting problems, and supports the idea that each of our limited number of causal precursors adds independent information to the forecast. Calculating the Bayesian information criterion (BIC), a metric similar to AIC, gives consistent results (see Table S3).

Moreover, we use the Niño-3.4 index to predict ISMR based on a simple linear regression model at 4-month lead time (see Figs. S2 and S3). For the same training and testing periods as for the CPC 1979–2016 dataset, this index does not provide any significant skill (see

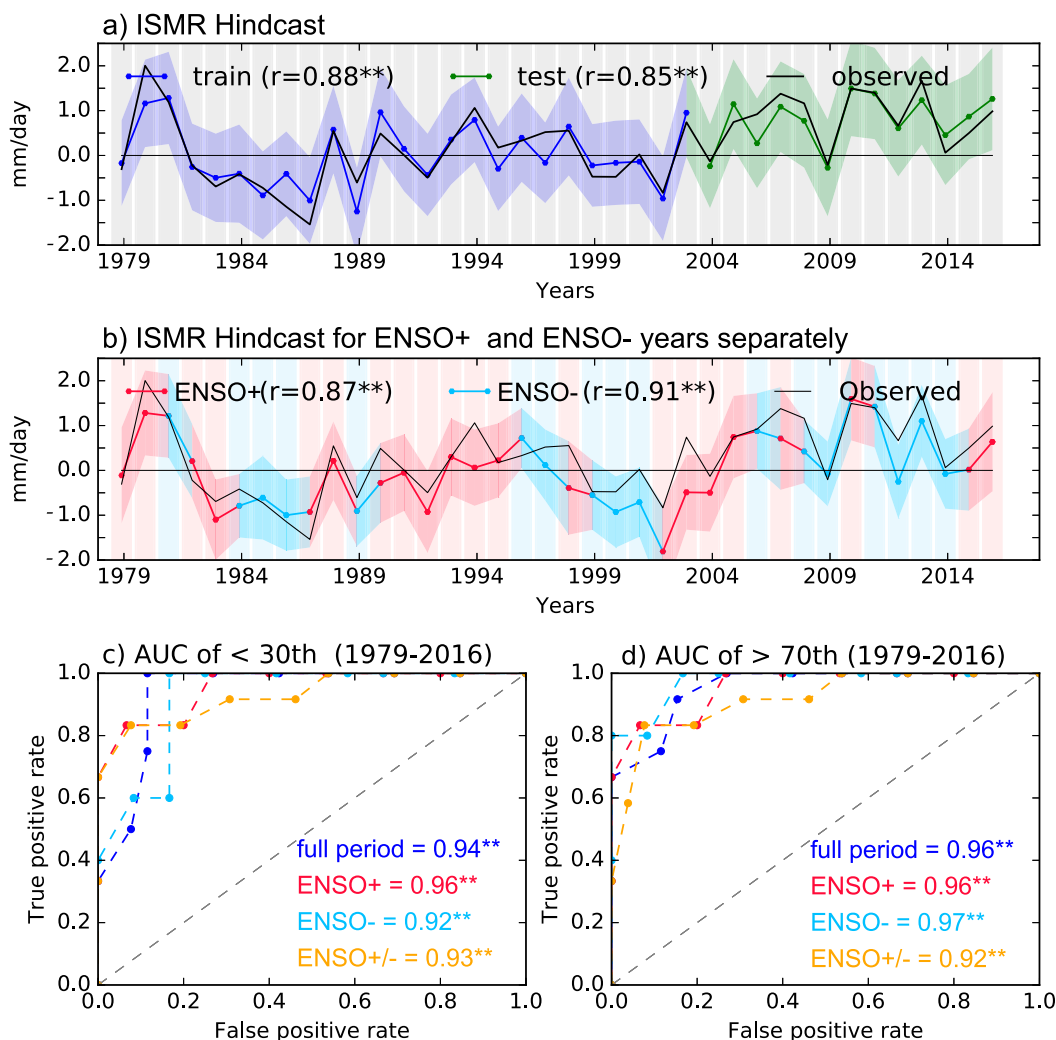


FIG. 5. Hindcasting ISMR. (a) CPC observations (solid black line) and hindcasts for training (1979–2003; solid blue lines) and testing (2004–16; solid green line) periods. Correlations between observed and hindcasted time series are also shown. (b) CPC observations (solid black line) and hindcasts for ENSO-positive (ENSO+; solid red line) and ENSO-negative (ENSO-; solid light blue line) years separately, with the resulting correlation values. (c) ROC curves for hindcasted events below the 30th percentile for the full 1979–2016 period (dashed blue line), for ENSO-positive years only (dashed red line), ENSO-negative years only (dashed light blue line), and for all years reconstructed using ENSO-positive and ENSO-negative years (dashed yellow line). (d) As in (c), but for events exceeding the 70th percentile. Double asterisks denote AUC values with  $p$  value < 0.05.

Figs. S2 and S3). Calculating the AIC for the combination of all five causal precursors and the Niño-3.4 index shows that the combination that gives the lowest AIC is the one that excludes the Niño-3.4 index (see Table S4).

### c. Forecast based on causal precursors

The causal precursors used to provide hindcasts in Fig. 5 are detected over the full period, implying that some information from the to-be-hindcasted year might enter the model. Therefore, we apply RG-CPD to provide an actual forecasting model that does not contain any information on the future. For this purpose, we use

RAJ together with ERA20C T2m and SLP fields for the 1901–2004 period and apply RG-CPD on a moving window of 30 years, and then forecast year 31, as would be the case in an operational forecast system (see Figs. 3b,c). We do not use CPC because of the shortness of the time series. A comparison between causal precursors obtained from CPC and RAJ datasets over the common period (1979–2004) and from the RAJ over the 1951–2004 period is shown in Fig. S4. Related hindcasts time series are shown in Fig. S5.

For each to-be-forecasted year, one forecast model is built based on all causal precursors identified in the

TABLE 1. Akaike information criterion (AIC). For the period 1979–2016, AIC values for the hindcast model are reported for individual causal precursors, combinations of four causal precursors, and for all five causal precursors in the first column. For ENSO-positive years, AIC values for the hindcast model are reported for individual causal precursors, all possible combinations of two causal precursors, and for all three causal precursors together in the second column. The third column is the same as the second column, but for ENSO-negative years.

1979–2016			ENSO-positive			ENSO-negative		
1 precursor	T2mIndones <sub>−4</sub>	58.7	1 precursor	T2mCAsia <sub>−5</sub>	42.7	1 precursor	T2mIndianOc <sub>−5</sub>	20.1
	T2mArctic <sub>−4</sub>	58.0		T2mAntarctic <sub>−4</sub>	42.6		T2mNAtl <sub>−5</sub>	14.1
	T2mArctic <sub>−5</sub>	56.7		T2mArctic <sub>−4</sub>	33.8	2 precursors	T2mIndianOc <sub>−5</sub> , T2mNAtl <sub>−5</sub>	12.1
	SLPPac <sub>−5</sub>	54.3	2 precursors	T2mCAsia <sub>−5</sub> , T2mAntarctic <sub>−4</sub>	38.9			
	T2mSAtl <sub>−4</sub>	54.2		T2mArctic <sub>−4</sub> , T2mCAsia <sub>−5</sub>	27.0			
4 precursors	SLPPac <sub>−5</sub> , T2mArctic <sub>−4</sub> , T2mIndones <sub>−4</sub> , T2mArctic <sub>−5</sub>	51.8		T2mArctic <sub>−4</sub> , T2mAntarctic <sub>−4</sub>	24.0			
	T2mSAtl <sub>−5</sub> , T2mArctic <sub>−4</sub> , T2mIndones <sub>−4</sub> , T2mArctic <sub>−5</sub>	43.5	3 precursors	T2mArctic <sub>−4</sub> , T2mCAsia <sub>−5</sub> , T2mAntarctic <sub>−4</sub>	21.9			
	SLPPac <sub>−5</sub> , T2mArctic <sub>−4</sub> , T2mArctic <sub>−5</sub> , T2mSAtl <sub>−4</sub>	36.2						
	SLPPac <sub>−5</sub> , T2mArctic <sub>−4</sub> , T2mIndones <sub>−4</sub> , T2mSAtl <sub>−4</sub>	34.6						
	T2mIndones <sub>−4</sub> , T2mArctic <sub>−5</sub> , SLPPac <sub>−5</sub> , T2mSAtl <sub>−4</sub>	32.7						
	SLPPac <sub>−5</sub> , T2mArctic <sub>−4</sub> , T2mIndones <sub>−4</sub> , T2mArctic <sub>−5</sub> , T2mSAtl <sub>−4</sub>	32.5						

five sets (see section 2 and Fig. 3c). Note that since the period 1901–1930 is used as a training period for the first 30-yr moving window, the first forecasted year is 1931.

Analyzing IMD predictions for the period 1988–2004 (3-month lead time) reveals low nonsignificant correlation with IMD observed ISMR ( $CC \sim 0.22$ ,  $p$  value  $> 0.1$ ; Fig. 1d), while AUC values for dry and wet events show no skill ( $AUC \sim 0.5$ ; Figs. 6c,d).

Here we show that building a forecast system based on causal precursors identified via RG-CPD provides some significant skill at different lead times. Forecasts obtained with forecast model at 4-month lead time (Fig. 6a) and 2-month lead time (Fig. 6b) for the period 1981–2004 are shown together with correlations and root mean squared errors relative to observations. Figures 6c and 6d show ROC curves and associated AUC values for dry events and wet events. The full forecast period 1931–2004 is shown in Fig. S6.

The forecasting skill shown in Fig. 6 drops as compared to the previously reported hindcast skill (Fig. 5 and Fig. S4), however, it remains significant. Correlation values between the observed and forecasted ISMR over the 1981–2004 period are  $CC = 0.37$  (4-month lead time,  $p$  value  $< 0.1$ ) and  $CC = 0.41$  (2-month lead time,

$p$  value  $< 0.05$ , see Figs. 6a,b). Our method shows significant, though modest, AUC values for both wet and dry events, and at both 4- and 2-month lead times. Specifically, the AUC values for 4-month lead time show significant skill with  $AUC = 0.73$  for dry events ( $p$  value  $< 0.1$ ) and  $AUC = 0.75$  for wet events ( $p$  value  $< 0.05$ ). AUC values for 2-month lead time show significant skill only for wet events, with  $AUC = 0.76$  ( $p$  value  $< 0.05$ ). Forecasting skills are summarized in Table 2.

Over the full 1931–2004 period (see Fig. S6), both AUC and correlation values are smaller in magnitude, nevertheless still significant at least for 2-month lead time with  $CC = 0.2$  ( $p$  value  $< 0.1$ ) and  $AUC = 0.64$  for wet events ( $p$  value  $< 0.05$ ). The drop in AUC and correlation scores is mainly due to the fact that the presented forecasting method fails in predicting ISMR amounts for the period 1957–80 (see Fig. S6). Contrarily, for the periods 1931–56 and 1981–2004 correlation values are about 0.4 and AUC values reach  $\sim 0.8$  (see Tables S5 and S6). A reason for this behavior may lie in the influence of the Pacific decadal oscillation (PDO) phase on the forecast performance: periods of good predictability correspond to positive PDO phases, while the 1957–80 period showing no predictability corresponds to a negative PDO phase (see Fig. S7).



TABLE 2. Summary of results. The correlation coefficients (CC) and AUC scores for actual forecasts obtained on different time periods and datasets are reported. IMD forecasting skill is also presented for comparison. AUC+ denotes AUC scores for events higher than the 70th percentile, and AUC− denotes AUC scores for events lower than the 30th percentile. Single (double) asterisk(s) denote AUC values with  $p$  value  $< 0.1$  (0.05).

RG-CPD based—4-month lead time				RG-CPD based—2-month lead time			
	CC	AUC−	AUC+		CC	AUC−	AUC+
1988–2004	0.35	0.68	0.85**	1988–2004	0.36	0.68	0.54
1981–2004	0.37*	0.73**	0.75**	1981–2004	0.41**	0.61	0.76**
1931–2004	0.15	0.54	0.62*	1931–2004	0.20*	0.55	0.64**
ENSO–RG-CPD based—4-month lead time				ENSO–RG-CPD based—2-month lead time			
1981–2004	0.03	0.55	0.63	1981–2004	0.36*	0.76**	0.69*
IMD							
	CC			AUC−		AUC+	
1988–2004	0.22			0.53		0.50	

shown in the form of frequency plots in Figs. S9 and S10, respectively. In these plots, the color indicates how many times a region has been used as an ISMR predictor. Frequency plots for different time periods (1933–56, 1957–80, and 1981–2004) underscore the nonstationarity of the precursors. Physical mechanisms that underlie the identified causal precursors are further discussed in the discussion section.

## 4. Discussion

### a. Hindcast and forecast skill and potential

Our study shows that a multiple linear regression model based on causal precursors from monthly SLP and T2m fields over the 1979–2016 period, provides a good hindcast of ISMR from CPC with a 4-month lead time (CC  $\sim 0.8$ , AUC  $\sim 0.95$ ). Considering the ENSO background state only slightly increases hindcast skills, nevertheless, it provides insightful information on the different mechanisms that are at play during different ENSO phases.

RG-CPD detects ISMR causal precursors, avoiding spurious correlations and identifying a small set of precursors: our precursor set (2–5 variables) is smaller than that used in the IMD model (5–9 variables), helping to limit overfitting of statistical regression models (DelSole and Shukla 2002; Wilks 2006). AIC values (Table 1) show that the combination of causal precursors chosen via RG-CPD is optimal when compared to smaller combinations of the same precursors. Thus, using all the detected precursors together actually improves the model, showing that higher correlation values are not an artifact of overfitting. Moreover, causal precursors detected with RG-CPD show a strong similarity with the set of precursors used for IMD operational forecasts (Rajeevan et al. 2007). Specifically, the IMD operational

forecasts use, among others, precursors from the central Pacific, western Indian Ocean, Scandinavia, eastern Eurasia, the Gulf of Mexico, and the North Atlantic, as described in the introduction section (Rajeevan et al. 2005). Our set of regions largely overlaps with IMD precursors: the most robust precursors across different time periods and datasets are the central/western Pacific causal precursors and the Arctic and Eurasian causal precursors, while the North Atlantic seems to be important in particular during ENSO-negative phases. This is encouraging, given that RG-CPD does not use any a priori preselection of the possible causal drivers and the causal precursors are detected among a set of more than one hundred precursor regions, all significantly correlated with ISMR. However, RG-CPD also detects new patterns such as causal precursors over the Southern Oceans and the set of precursors can easily be updated with new data becoming available (see Figs. S9 and S10).

Actual forecasts skill, that is, with no information from the future entering the RG-CPD process, decreases with respect to hindcasts, likely due to the strong nonstationarity of the precursors. Still, our method produces significant correlation (CC  $\sim 0.4$ ,  $p$  value  $< 0.05$ ) and AUC  $\sim 0.7$ – $0.8$  ( $p$  value  $< 0.05$ ) values at both 2- and 4-month lead time over the 1981–2004 period, outperforming IMD forecasts when compared over the same period (Fig. 6).

RG-CPD provides a complementary method to ANN (Singh and Borah 2013; Singh 2018). While both methods provide objective and automatized forecasts, causal precursors detected via RG-CPD can be interpreted from a physical point of view, while ANN extracts the predictability from the ISMR time series itself. RG-CPD provides forecasts with lower skill when compared to Singh (2018), but at 2-month lead time (instead of 1 month).



Moreover, with RG-CPD it is possible to trace back all calculation that lead to discard or retain any of the causal precursors. Finally, RG-CPD makes it straightforward to develop forecasts for specific regions and time scales of interest and can be easily applied to provide real-time seasonal ISMR forecasts.

### *b. Physical interpretation of causal precursors*

Causal precursors need to be considered together with theoretical understanding of the physical mechanisms which lie behind statistical causality. In general, many of the detected causal precursors show similar patterns as IMD (Rajeevan et al. 2005) and as Saha et al. (2016), in particular when a 2-month lead time is considered (see Figs. S1 and S9). The physical interpretation of these patterns has been very difficult (Rajeevan 2001; Gadgil et al. 2005; Rajeevan et al. 2007). While a detailed analysis into that is outside the scope of this paper, here we discuss previously proposed physical hypotheses that help explaining the detected precursor regions.

To propagate into summer, winter and early spring phenomena need to be governed by slowly changing components of the land–ocean system with a long memory, such as SST, snow cover, or soil moisture. Tropical SSTs are crucial in modulating the intensity of the ISMR (Wang et al. 2005). Accordingly, most of the detected causal precursors are located over the oceans. SLP<sup>Pacific</sup> <sub>$\tau=-5$</sub>  is the largest detected region in the SLP field over the equatorial Pacific and Indian Ocean. Despite interdecadal oscillation, both models and observations show that the location and intensity of the Walker circulation and ENSO are essential in shaping the ISMR (Krishnan et al. 1998; Krishna Kumar et al. 1999; Krishnamurthy and Goswami 2000; Mujumdar et al. 2007). Depending on the specific period, the shape and geographical position of the causal precursors coming from the Pacific Ocean can change. Still, a central Pacific precursor is almost always present in the set of causal precursors and is also used by IMD forecasts (Rajeevan et al. 2007).

While soil moisture does not seem to have a strong direct influence on the ISMR (Shinoda 2001), the snow–monsoon mechanism supports the hypothesis that Eurasian snow cover conditions in winter to early spring might influence the onset and the intensity of the ISMR (Hahn and Shukla 1976; Bamzai and Shukla 1999; Dash et al. 2005, 2006) and have been often used by IMD forecasts (Rajeevan et al. 2004). T2m<sup>CAsia</sup> <sub>$\tau=-5$</sub>  is negatively correlated with ISMR, and corresponds to the East Asia SLP region also used by IMD for their operational forecasts (Rajeevan et al. 2007; Kumar 2012). Colder temperatures are positively correlated with higher snow cover (Bamzai and Shukla 1999). Although the signal we find is shifted further westward, T2m<sup>CAsia</sup> <sub>$\tau=-5$</sub>  supports prior

evidence that the strongest snow cover–ISMR link originates from eastern Eurasia (Peings and Douville 2010).

T2m<sup>Arctic</sup> <sub>$\tau=-5$</sub>  and T2m<sup>Arctic</sup> <sub>$\tau=-4$</sub>  show a positive correlation between ISMR and Arctic temperatures in particular during ENSO-positive years supporting prior studies that link higher temperatures over the Northern Hemisphere during January with enhanced ISMR (Sikka 1980; Verma et al. 1985). Similarly, temperature over Eurasia in February–March is linked with enhanced ISMR (Rajeevan et al. 1998). In our study, T2m anomalies over Eurasia, north of 60°N, show significant positive correlation with ISMR (T2m<sup>Arctic</sup> <sub>$\tau=-5$</sub> ), while central Asia around 40°N (T2m<sup>CAsia</sup> <sub>$\tau=-5$</sub> ) sees a reversed relationship (Rajeevan et al. 1998). Similarly, IMD uses temperature anomalies from stations over Scandinavia (Rajeevan et al. 2007). The land–ocean temperature contrast is a well-known driver of the ISM circulation. Thus, the relationship between higher temperatures over the Arctic in winter and enhanced ISMR could be linked to the enhanced thermal contrast between the Eurasian land mass and the Indian Ocean in summer. The snow–monsoon mechanisms may further support these findings: positive temperature anomalies in the Arctic during winter could be linked to negative snow cover anomalies and thus to increased ISMR. Additionally, some studies show that the correlation between the ISMR and Eurasian snow cover is particularly strong in the northernmost regions (Mamgain et al. 2010). A second option could be that the Arctic region responds to signals coming from the tropical belt and is linked to ENSO, which is stronger in winter. The detected causal relationship with the ISMR might be an artifact of missing relevant variables or time scales in our analysis. Note again that the causal precursors identified are only *causal relative to the set of analyzed precursors*.

T2m<sup>NAtlantic</sup> <sub>$\tau=-4$</sub>  is also used as a precursor by IMD (Rajeevan et al. 2007). In this case, the physical linkage is less clear (Rajeevan et al. 2005), nevertheless IMD identifies some predictability for the ISMR from SST in the tropical North Atlantic and SLP over the Azores High region (likely linked to the NAO) (Rajeevan et al. 2007, 2005). Previous evidence suggests that the ISMR and the Atlantic Niño share an inverse relationship during summer months (Wang et al. 2009; Yadav et al. 2018). The Atlantic Niño significantly influences a dipole pattern of rainfall in the northern parts of India and the positive phase of the Atlantic Niño results in a local tropospheric warming, which intensifies the intertropical convergence zone over the equatorial east Atlantic and west Africa and excites an anticyclonic anomaly over India (Wang et al. 2009; Kucharski and Joshi 2017; Yadav et al. 2018). During March and April, a positive correlation is found between Atlantic tropical SST and the ISMR, as shown in our study (Varikoden and Babu 2015).

Moreover, models show that temperatures over the North Atlantic give useful skill in predicting temperatures over eastern Eurasia in summer at yearly time scale (Monerie et al. 2017). Keeping in mind the link between temperatures over Eurasia and the ISMR, this could further explain the link between North Atlantic T2m in winter and the ISMR.

Additional evidence for a link between the North Atlantic and the ISMR is provided also by the relationship detected between the Atlantic multidecadal oscillation (AMO) and multidecadal variability of the Indian summer monsoon rainfall (Goswami et al. 2006). The AMO weakens the meridional gradient of tropospheric temperature by setting up some negative temperature anomaly over Eurasia during boreal late summer/autumn, which results in an early withdrawal of the south west monsoon and decreased ISMR (Goswami et al. 2006). In May, OLR in the North Atlantic negatively correlates with the ISMR, further supporting the link between the NAO and the ISMR (Srivastava et al. 2002). The  $T2m_{T=-5}^{SWIndianOc}$  might be linked to higher SST over the southern Indian Ocean, also used by IMD, which during February and March enhance evaporation and fuel stronger ISMR (Rajeevan et al. 2002; Terray et al. 2003; Thapliyal and Rajeevan 2003).

Robust causal precursors are identified in the southern ocean, in particular in the South Atlantic (see Fig. S4). The South Atlantic anticyclone is a characteristic feature of the austral winter climatology and has been linked to the West African and Asian summer monsoon circulations during summer (Richter et al. 2008). In this study, the monsoon influences the South Atlantic anticyclone: the convergence centers associated with monsoon activity in the Northern Hemisphere affect large-scale subsidence over the South Atlantic (Richter et al. 2008; Sun et al. 2017). Here, we hypothesize that the opposite link holds as well.

The dependence of ISM precursors on the ENSO background state can help to better understand these underlying physical mechanisms. During ENSO-positive years, the predictability of ISMR originates from high latitudes (in both hemispheres), while during ENSO-negative years the predictability predominantly originates from the tropical belt. A possible explanation is that during ENSO-positive years the weakening of easterly trades over the tropical Pacific is accompanied by a pronounced extratropical circulation response to tropical convective anomalies. This creates patterns of alternating anomalous low- and high-pressure systems in an area extending from the tropics to midlatitudes (Lau and Lim 1984). This is consistent with a Rossby wave dispersion mechanism arising from interactions between anomalies of tropical convection and large-scale circulation (Hoskins and Karoly 1981). This situation is

reversed during ENSO-negative periods, when the atmospheric circulation response is confined to a much narrower tropical belt due to strong easterlies.

The ENSO phase might also affect the ISMR via its effect on the winter European climate (Brönnimann 2007). The ENSO-positive effect on the European climate is similar to that of a negative NAO phase and as previously discussed, NAO affects the ISMR via changes in the atmospheric circulation over Eurasia and related snow cover patterns (Dugam et al. 1997; Rajeevan 2002; Goswami et al. 2006).

## 5. Conclusions

Our study shows that a novel causal discovery algorithm coupled with a response-guided detection step can identify causal precursors of global atmospheric fields for seasonal all-India summer monsoon rainfall (ISMR). The Response-Guided Causal Precursor Detection (RG-CPD) algorithm identifies causal precursors without a priori assumptions, except for the selection of variables and prediction lead times. The global sea level pressure and 2-m temperatures are used in this study for the inferences of causal precursors, and a statistical hindcasting model is developed for predicting the seasonal ISMR at long (4-month) lead times using these precursors. The detected causal precursors provide appreciable hindcast skills with CC  $\sim 0.8$  and AUC scores  $\sim 0.9$ , and also are generally consistent with the variables used by the India Meteorological Department (IMD) for operational forecasting. Further accounting for the El Niño–Southern Oscillation (ENSO) background state in the analysis sheds additional insight into physical pathways during different ENSO phases.

Using causal precursors in a true operational mode without knowledge of any future years gives significant forecast skill at both shorter (2-month) and longer (4-month) lead times. While the forecast skill in an operational mode is higher than IMD skill, it is substantially lower than that of the hindcasts. Our forecast model yields significant CC  $\sim 0.4$  and AUC  $\sim 0.7$  for the period 1981–2004 and CC  $\sim 0.2$  and AUC  $\sim 0.6$  over the longer period 1931–2004. One of the advantages of using the RG-CPD procedure is that it detects a smaller set of variables which reduces the risk of overfitting, as confirmed with the Akaike Information Criterion analysis. In addition, since the precursors are nonstationary, this objective methodology provides a fast and automatized procedure to update precursors by itself in association with changing or emerging new patterns.

This study clearly provides evidences that causal discovery tools such as RG-CPD can significantly improve the physical understanding of far-away or remote drivers

influencing the ISMR at different time scales and also under different background conditions. We also conclude that this promising automatized technique can provide superior skill in seasonal ISMR forecasts at 2–4-month lead time, allowing for better strategic socioeconomic planning for the Indian subcontinent.

**Acknowledgments.** We thank ECMWF and NCEP for making the ERA-Interim and CPC data available. We also thank the anonymous reviewers for the helpful suggestions. G.D.C., D.C., and R.V.D. acknowledge cofunding from the German Federal Ministry of Education and Research (BMBF Grant 01LP1611A) under the auspices of the Belmont Forum and JPI-Climate project, GOTHAM. R.V.D. and M.K. acknowledge cofunding from the German Federal Ministry of Education and Research (BMBF Grants 01LN1306A-COSY and 01LN1304A-SACREX). This work was partially supported by the European Union's Horizon 2020 MSCA programme under Grant 704585 (PROCEED MSCA project; <http://projects.knmi.nl/proceed/>) and research and innovation programme under Grant 776868 (SECLI-FIRM project; <http://www.secli-firm.eu/>). Code for the causal discovery algorithm is freely available as part of the Tigramite Python software package at <https://github.com/jakobrunge/tigramite>.

## REFERENCES

- Bamzai, A. S., and J. Shukla, 1999: Relation between Eurasian snow cover, snow depth, and the Indian summer monsoon: An observational study. *J. Climate*, **12**, 3117–3132, [https://doi.org/10.1175/1520-0442\(1999\)012<3117:RBESCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<3117:RBESCS>2.0.CO;2).
- Bello, G. A., M. Angus, N. Pedemane, J. K. Harlalka, F. H. M. Semazzi, V. Kumar, and N. F. Samatova, 2015: Response-guided community detection: Application to climate index discovery. *Machine Learning and Knowledge Discovery in Databases*, A. Appice et al., Eds., Springer, 736–751.
- Bhanu Kumar, O. S. R. U., 1988: Eurasian snow cover and seasonal forecast of Indian summer monsoon rainfall. *Hydrol. Sci. J.*, **33**, 515–525, <https://doi.org/10.1080/02626668809491278>.
- Brönnimann, S., 2007: Impact of El Niño–Southern on European climate. *Rev. Geophys.*, **45**, <https://doi.org/10.1029/2006RG000199>.
- Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.*, **113**, D04110, <https://doi.org/10.1029/2007JD009132>.
- Cherchi, A., and A. Navarra, 2013: Influence of ENSO and of the Indian Ocean Dipole on the Indian summer monsoon variability. *Climate Dyn.*, **41**, 81–103, <https://doi.org/10.1007/s00382-012-1602-y>.
- Dash, S. K., G. P. Singh, A. D. Vernekar, and M. S. Shekhar, 2004: A study on the number of snow days over Eurasia, Indian rainfall and seasonal circulations. *Meteor. Atmos. Phys.*, **86**, 1–13, <https://doi.org/10.1007/s00703-003-0016-0>.
- , —, M. S. Shekhar, and A. D. Vernekar, 2005: Response of the Indian summer monsoon circulation and rainfall to seasonal snow depth anomaly over Eurasia. *Climate Dyn.*, **24**, 1–10, <https://doi.org/10.1007/s00382-004-0448-3>.
- , P. Parth Sathi, and S. K. Panda, 2006: A study on the effect of Eurasian snow on the summer monsoon circulation and rainfall using a spectral GCM. *Int. J. Climatol.*, **26**, 1017–1025, <https://doi.org/10.1002/joc.1299>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- DelSole, T., and J. Shukla, 2002: Linear prediction of Indian monsoon rainfall. *J. Climate*, **15**, 3645–3658, [https://doi.org/10.1175/1520-0442\(2002\)015<3645:LPOIMR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3645:LPOIMR>2.0.CO;2).
- , and —, 2009: Artificial skill due to predictor screening. *J. Climate*, **22**, 331–345, <https://doi.org/10.1175/2008JCLI2414.1>.
- Dugam, S. S., S. B. Kakade, and R. K. Verma, 1997: Interannual and long-term variability in the North Atlantic Oscillation and Indian Summer monsoon rainfall. *Theor. Appl. Climatol.*, **58**, 21–29, <https://doi.org/10.1007/BF00867429>.
- Fasullo, J., 2004: A stratified diagnosis of the Indian monsoon–Eurasian snow cover relationship. *J. Climate*, **17**, 1110–1122, [https://doi.org/10.1175/1520-0442\(2004\)017<1110:ASDOTI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1110:ASDOTI>2.0.CO;2).
- Ferranti, L., and F. Molteni, 1999: Ensemble simulations of Eurasian snow-depth anomalies and their influence on the summer Asian monsoon. *Quart. J. Roy. Meteor. Soc.*, **125**, 2597–2610, <https://doi.org/10.1002/qj.49712555913>.
- Gadgil, S., and K. Rupa Kumar, 2006: The Asian Monsoon – Agriculture and economy. *The Asian Monsoon*, B. Wang, Ed., Springer, 651–683.
- , M. Rajeevan, and R. Nanjundiah, 2005: Monsoon prediction – Why yet another failure? *Curr. Sci.*, **88**, 1389–1400, <https://www.jstor.org/stable/24110705>.
- Goswami, B. N., and P. K. Xavier, 2005: Dynamics of “internal” interannual variability of the Indian summer monsoon in a GCM. *J. Geophys. Res.*, **110**, D24104, <https://doi.org/10.1029/2005JD006042>.
- , M. S. Madhusoodanan, C. P. Neema, and D. Sengupta, 2006: A physical mechanism for North Atlantic SST influence on the Indian summer monsoon. *Geophys. Res. Lett.*, **33**, L02706, <https://doi.org/10.1029/2005GL024803>.
- Hahn, D. G., and J. Shukla, 1976: An apparent relationship between Eurasian snow cover and Indian monsoon rainfall. *J. Atmos. Sci.*, **33**, 2461–2462, [https://doi.org/10.1175/1520-0469\(1976\)033<2461:AARBES>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<2461:AARBES>2.0.CO;2).
- Hastenrath, S., and L. Greischar, 1993: Changing predictability of Indian monsoon rainfall anomalies? *Proc. Indian Acad. Sci., Earth Planet. Sci.*, **102**, 35–47, <https://doi.org/10.1007/BF02839181>.
- Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, [https://doi.org/10.1175/1520-0469\(1981\)038<1179:TSLROA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2).
- Jain, S., A. A. Scaife, and A. K. Mitra, 2019: Skill of Indian summer monsoon rainfall prediction in multiple seasonal prediction systems. *Climate Dyn.*, **52**, 5291–5301, <https://doi.org/10.1007/s00382-018-4449-z>.
- Kretschmer, M., D. Coumou, J. F. Donges, and J. Runge, 2016: Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *J. Climate*, **29**, 4069–4081, <https://doi.org/10.1175/JCLI-D-15-0654.1>.
- , J. Runge, and D. Coumou, 2017: Early prediction of extreme stratospheric polar vortex states based on causal precursors.

- Geophys. Res. Lett.*, **44**, 8592–8600, <https://doi.org/10.1002/2017GL074696>.
- Kripalani, R. H., A. Kulkarni, and S. S. Sabade, 2001: El Niño–Southern Oscillation, Eurasian snow cover and the Indian monsoon rainfall. *Indian Natl. Sci. Acad.*, **67**, 361–368.
- Krishna Kumar, K., B. Rajagopalan, and M. Cane, 1999: On the weakening relationship between the Indian monsoon and ENSO. *Science*, **284**, 2156–2159, <https://doi.org/10.1126/science.284.5423.2156>.
- Krishnamurthy, V., and B. N. Goswami, 2000: Indian monsoon–ENSO relationship on interdecadal timescale. *J. Climate*, **13**, 579–595, [https://doi.org/10.1175/1520-0442\(2000\)013<0579:IMEROI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0579:IMEROI>2.0.CO;2).
- Krishnan, R., C. Venkatesan, and R. N. Keshavamurthy, 1998: Dynamics of upper tropospheric stationary wave anomalies induced by ENSO during the northern summer: A GCM study. *Proc. Indian Acad. Sci., Earth Planet. Sci.*, **107**, 65–90, <https://doi.org/10.1007/BF02842261>.
- Kucharski, F., and M. K. Joshi, 2017: Influence of tropical South Atlantic sea-surface temperatures on the Indian summer monsoon in CMIP5 models. *Quart. J. Roy. Meteor. Soc.*, **143**, 1351–1363, <https://doi.org/10.1002/qj.3009>.
- Kumar, A., 2012: Statistical models for long-range forecasting of southwest monsoon rainfall over India using step wise regression and neural network. *Atmos. Climate Sci.*, **2**, 322–336, <https://doi.org/10.4236/acs.2012.23029>.
- Lau, K.-M., and H. Lim, 1984: On the dynamics of equatorial forcing of climate teleconnections. *J. Atmos. Sci.*, **41**, 161–176, [https://doi.org/10.1175/1520-0469\(1984\)041<0161:OTDOEF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041<0161:OTDOEF>2.0.CO;2).
- Mamgain, A., S. K. Dash, and P. P. Sarthi, 2010: Characteristics of Eurasian snow depth with respect to Indian summer monsoon rainfall. *Meteor. Atmos. Phys.*, **110**, 71–83, <https://doi.org/10.1007/s00703-010-0100-1>.
- Monerie, P. A., J. Robson, B. Dong, and N. Dunstone, 2017: A role of the Atlantic Ocean in predicting summer surface air temperature over North East Asia? *Climate Dyn.*, **51**, 473–491, <https://doi.org/10.1007/s00382-017-3935-z>.
- Mujumdar, M., V. Kumar, and R. Krishnan, 2007: The Indian summer monsoon drought of 2002 and its linkage with tropical convective activity over northwest Pacific. *Climate Dyn.*, **28**, 743–758, <https://doi.org/10.1007/s00382-006-0208-7>.
- Peings, Y., and H. Douville, 2010: Influence of the Eurasian snow cover on the Indian summer monsoon variability in observed climatologies and CMIP3 simulations. *Climate Dyn.*, **34**, 643–660, <https://doi.org/10.1007/s00382-009-0565-0>.
- Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29**, 4083–4097, <https://doi.org/10.1175/JCLI-D-15-0556.1>.
- Rajeevan, M., 2001: Prediction of Indian summer monsoon: Status, problems and prospects. *Curr. Sci.*, **81**, 1451–1457.
- , 2002: Winter surface pressure anomalies over Eurasia and Indian summer monsoon. *Geophys. Res. Lett.*, **29**, 1454, <https://doi.org/10.1029/2001GL014363>.
- , D. S. Pai, and V. Thapliyal, 1998: Spatial and temporal relationships between global land surface air temperature anomalies and Indian summer monsoon rainfall. *Meteor. Atmos. Phys.*, **66**, 157–171, <https://doi.org/10.1007/BF01026631>.
- , —, and —, 2002: Predictive relationships between Indian Ocean sea surface temperatures and Indian summer monsoon rainfall. *Mausam*, **53**, 337–348.
- , —, S. K. Dikshit, and R. R. Kelkar, 2004: IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. *Curr. Sci.*, **86**, 422–431.
- , —, and R. A. Kumar, 2005: New statistical models for long range forecasting of south west monsoon rainfall over India. NCC Research Rep. 1, 51 pp., [http://www.imdpune.gov.in/Clim\\_Pred\\_LRF\\_New/Reports/NCCResearchReports/research\\_report\\_1.pdf](http://www.imdpune.gov.in/Clim_Pred_LRF_New/Reports/NCCResearchReports/research_report_1.pdf).
- , —, R. Anil Kumar, and B. Lal, 2007: New statistical models for long-range forecasting of southwest monsoon rainfall over India. *Climate Dyn.*, **28**, 813–828, <https://doi.org/10.1007/s00382-006-0197-6>.
- , J. Bhatte, and A. K. Jaswal, 2008: Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett.*, **35**, L18707, <https://doi.org/10.1029/2008GL035143>.
- Richter, I., C. R. Mechoso, and A. W. Robertson, 2008: What determines the position and intensity of the South Atlantic anticyclone in austral winter?—An AGCM study. *J. Climate*, **21**, 214–229, <https://doi.org/10.1175/2007JCLI1802.1>.
- Robock, A., M. Mu, K. Vinnikov, and D. Robinson, 2003: Land surface conditions over Eurasia and Indian summer monsoon rainfall. *J. Geophys. Res.*, **108**, 4131, <https://doi.org/10.1029/2002JD002286>.
- Runge, J., 2018: Causal network reconstruction from time-series: From theoretical assumptions to practical estimation. *Chaos*, **28**, 075310, <https://doi.org/10.1063/1.5025050>.
- , J. Heitzig, N. Marwan, and J. Kurths, 2012: Quantifying causal coupling strength: A lag-specific measure for multivariate time-series related to transfer entropy. *Phys. Rev. E*, **86**, 061121, <https://doi.org/10.1103/PhysRevE.86.061121>.
- , V. Petoukhov, and J. Kurths, 2014: Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *J. Climate*, **27**, 720–739, <https://doi.org/10.1175/JCLI-D-13-00159.1>.
- , and Coauthors, 2015a: Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Commun.*, **6**, 9502, <https://doi.org/10.1038/ncomms9502>.
- , R. V. Donner, and J. Kurths, 2015b: Optimal model-free prediction from multivariate time-series. *Phys. Rev. E*, **91**, 052909, <https://doi.org/10.1103/PhysRevE.91.052909>.
- Saha, K., F. Sanders, and J. Shukla, 1981: Westward propagating predecessors of monsoon depressions. *Mon. Wea. Rev.*, **109**, 330–343, [https://doi.org/10.1175/1520-0493\(1981\)109<0330:WPPOMD>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0330:WPPOMD>2.0.CO;2).
- Saha, M., P. Mitra, and R. S. Nanjundiah, 2016: Autoencoder-based identification of predictors of Indian monsoon. *Meteor. Atmos. Phys.*, **128**, 613–628, <https://doi.org/10.1007/s00703-016-0431-7>.
- Sahai, A. K., M. K. Soman, and V. Satyan, 2000: All India summer monsoon rainfall prediction using an artificial neural network. *Climate Dyn.*, **16**, 291–302, <https://doi.org/10.1007/s003820050328>.
- Shinoda, M., 2001: Climate memory of snow mass as soil moisture over central Eurasia. *J. Geophys. Res.*, **106**, 33 393–33 403, <https://doi.org/10.1029/2001JD000525>.
- Shukla, J., and D. A. Mooley, 1987: Empirical prediction of the summer monsoon rainfall over India. *Mon. Wea. Rev.*, **115**, 695–703, [https://doi.org/10.1175/1520-0493\(1987\)115<0695:EPOTSM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0695:EPOTSM>2.0.CO;2).
- Shukla, R. P., K. C. Tripathi, A. C. Pandey, and I. M. L. Das, 2011: Prediction of Indian summer monsoon rainfall using Niño indices: A neural network approach. *Atmos. Res.*, **102**, 99–109, <https://doi.org/10.1016/j.atmosres.2011.06.013>.



- Sikka, D. R., 1980: Some aspects of the large-scale fluctuations of summer monsoon rainfall over India in relation to fluctuations in the planetary and regional scale circulations. *Proc. Indian Acad. Sci., Earth Planet. Sci.*, **89**, 179–195, <https://doi.org/10.1007/BF02913749>.
- Singh, P., 2018: Indian summer monsoon rainfall (ISMR) forecasting using time-series data: A fuzzy-entropy-neuro based expert system. *Geosci. Front.*, **9**, 1243–1257, <https://doi.org/10.1016/j.gsf.2017.07.011>.
- , and B. Borah, 2013: Indian summer monsoon rainfall prediction using artificial neural network. *Stochastic Environ. Res. Risk Assess.*, **27**, 1585–1599, <https://doi.org/10.1007/s00477-013-0695-0>.
- Spirtes, P., C. Glymour, and R. Scheines, 2000: *Causation, Prediction, and Search*. MIT Press, 568 pp.
- Srivastava, A. K., M. Rajeevan, and R. Kulkarni, 2002: Teleconnection of OLR and SST anomalies over Atlantic Ocean with Indian summer monsoon. *Geophys. Res. Lett.*, **29**, 1284, <https://doi.org/10.1029/2001gl013837>.
- Sun, X., K. H. Cook, and E. K. Vizy, 2017: The South Atlantic subtropical high: Climatology and interannual variability. *J. Climate*, **30**, 3279–3296, <https://doi.org/10.1175/JCLI-D-16-0705.1>.
- Terray, P., P. Delecluse, S. Labattu, and L. Terray, 2003: Sea surface temperature associations with the late Indian summer monsoon. *Climate Dyn.*, **21**, 593–618, <https://doi.org/10.1007/s00382-003-0354-0>.
- Thapliyal, V., and M. Rajeevan, 2003: Updated operational models for long-range forecasts of Indian summer monsoon rainfall. *Mausam*, **54**, 495–504.
- Turner, A. G., and J. M. Slingo, 2011: Using idealized snow forcing to test teleconnections with the Indian summer monsoon in the Hadley Centre GCM. *Climate Dyn.*, **36**, 1717–1735, <https://doi.org/10.1007/s00382-010-0805-3>.
- Varikoden, H., and C. A. Babu, 2015: Indian summer monsoon rainfall and its relation with SST in the equatorial Atlantic and Pacific Oceans. *Int. J. Climatol.*, **35**, 1192–1200, <https://doi.org/10.1002/joc.4056>.
- Verma, R. K., K. Subramaniam, and S. S. Dugam, 1985: Inter-annual and long-term variability of the summer monsoon and its possible link with northern hemispheric surface air temperature. *Proc. Indian Acad. Sci., Earth Planet. Sci.*, **94**, 187–198, <https://doi.org/10.1007/BF02839197>.
- Wang, B., Q. Ding, X. Fu, I. S. Kang, K. Jin, J. Shukla, and F. Doblas-Reyes, 2005: Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophys. Res. Lett.*, **32**, L15711, <https://doi.org/10.1029/2005GL023769>.
- Wang, C., F. Kucharski, R. Barimalala, and A. Bracco, 2009: Teleconnections of the tropical Atlantic to the tropical Indian and Pacific Oceans: A review of recent findings. *Meteor. Z.*, **18**, 445–454, <https://doi.org/10.1127/0941-2948/2009/0394>.
- Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, <https://doi.org/10.1098/rsif.2013.1162>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.
- Willink, D., V. Khan, and R. V. Donner, 2017: Improved one-month lead time forecasting of the SPI over Russia with pressure covariates based on the SL–AV model. *Quart. J. Roy. Meteor. Soc.*, **143**, 2636–2649, <https://doi.org/10.1002/qj.3114>.
- Yadav, R. K., G. Srinivas, and J. S. Chowdary, 2018: Atlantic Niño modulation of the Indian summer monsoon through Asian jet. *npj Climate Atmos. Sci.*, **1**, 23, <https://doi.org/10.1038/s41612-018-0029-5>.